Relationship between GDP per Capita and Income Inequality

Mark Kimura Smith

Georgia Institute of Technology

Abstract:

This paper examines the relationship between GDP per capita and income inequality with the hypothesis that there is a negative relation between an areas income inequality compared to its GDP. The GINI coefficient was utilized to measure income inequality along with regression models to analyze other relevant factors. Cross-country data was used to find how various independent variables would affect a countries income inequality (GINI index/dependent variable).

I. Introduction

Income inequality has been a very controversial topic both economically and politically with arguments surrounding the overall impact it has on society. According to Emmanuel Saez, an economist at UC Berkley, American individuals in the top 10% average more than nine times the income than those in the bottom 90%. Due to drastic income disparities that were present in the early 1900s, reforms were created by policymakers such as increasing taxes on the wealthy and increasing unionization for those at the bottom, but these began to erode in the 1970s, leading to even an even bigger gap today.

Although the relationship between GDP and inequality has been heavily researched for many years, it is a topic that still has no clear resolution. When it comes to the argument for why greater inequality is beneficial, a main one is that it creates an incentive to work more. For example, if those with higher levels of education are more productive, differences in income will encourage more individuals to attain higher education, increasing overall productivity. Another is that it will lead to more economic growth through investment, with the trend that those with higher incomes are more likely to save and invest than their counterparts.

On the other hand, there are those who argue that greater inequality brings up moral issues because it decreases opportunities for minorities and groups at a disadvantage. When traditionally looking at this issue, many may believe that the main issue involves morals, but there are economic effects too. If marginalized groups are set back by the income inequality, there is a decrease in social mobility, which in turn will decrease economic growth and the investment into human capital of low earning individuals.

Overall, this paper will aim to look at these conflicting viewpoints by analyzing what changes in GDP will have on the GINI coefficient. Although the main independent variable is the GDP, I will also be controlling for other factors that could affect income inequality, such as unemployment and savings.

II. Literature Review

Panel and cross-sectional data taken from multiple countries can be utilized to see how different factors effect GDP and this was done by Barro (2000). Barro begins his analysis by identifying some macroeconomic consequences of income inequality, which includes things such as the political economy, credit-market imperfections, and savings rates. He then draws upon a panel of roughly 100 different countries between the years 1960 and 1995 and uses regression models such as investments, terms of trade, democracy index, government consumption, among other variables to determine the growth rate within these countries. A large portion of the analysis used in this study uses the Gini coefficient to measure the income inequality, which relates to the Lorenz curve that graphs the cumulated income shares and population shares against each other. One of the main conclusions that Barro arrived to after conducting his research is that income inequality tends to slow down growth in poor/developing countries, while having the inverse effect on rich/developed countries. More specifically, he concluded that growth tends to fall with inequality when per capita GDP is below $2000 (1985 dollar value) and rises with inequality when per capita GDP is above $2000.

When it comes to comparing GDP and income inequality, the Kuznet Curve developed by Simon Kuznet (1955) is drawn upon in many literatures, including Barros'. His work utilized data from 3 countries; the United States, England, and Germany, with the main overview being that income inequality increases as a country is developing, specifically from a rural to urban population, and inversely decreases when the modern structure become prevalent. This relation is shown in the Kuznet Curve, which illustrates an inverted U-shape for the relation between income inequality and per capita income. Initially, the curve looks at the relation between inequality and level of output, which created conflict between if it was detrimental or necessary, so current literature focuses more on the relationship between growth and inequality. Kuznet stressed 3 aspects of his findings, the first being that the data he used was for income before direct taxes and excludes government contributions. Secondly, he states that reductions in percentage of inequality was accompanied by drastic rises in real income per capita, meaning that countries that are classified as developed are experiencing increase in income per capita if there is no conflict such as war. Lastly, he brings up the question of weather distributions by annual incomes properly reflect trends by secular income. This primarily speculates that long term average incomes may show a smaller reduction in inequality compared to annual incomes.

Ikemoto and Uehara (2000) illustrate the Kuznet curve more specifically by looking at its relation to income inequality in Thailand. In their literature, they hypothesized that Thailand, which saw rapid economic growth in the 1980s and with the industrial sector absorbing the underemployed rural labor force in the 1990s, would soon seen a decrease in income inequality. After conducting an analysis of the GINI coefficient to poverty across Thailand, they were surprised to find that after the country had already passed the Kuznet Curve, the income inequality increased again, meaning that the U-shaped curve was more like a N-shaped curve. Ikemoto and Uehara then revisited the Kuznets' hypothesis which is based off a transition from an agricultural economy to an industrial one, meaning it is only supposed to happen a single time during economic development. They then came to the conclusion that changes in new high productivity industry could affect the Kuznets Curve, and it should not be limited to only a change from a agricultural to industrial economy.

Banerjee (2003) looks at existing literatures on the topic of income inequality and its effect on economic growth, analyzing why different approaches lead to mixed results. When using OLS regressions using one cross section, it is typically found that there is a positive relationship between inequality and growth, while the fixed effect approach produces a negative relation between changes in inequality and changes in growth rate. Banerjee believes that it may not be possible to interpret the evidence in these literatures casually, and that variations in inequality could likely be credited to a range of unobservable factors associated with growth.

The contribution that I will add to the already extensive literature is that because many different conclusions have been reached when it comes to this topic, I will test these by comparing it to my findings seeing if there is a match with a previous one or perhaps no match at all. I will also be taking a deeper look at the Kuznets Curve across a more extensive set of cross-country data instead of just one. I will also use a more macroeconomic view when comparing GDP per capita and income inequality so that there will be a broader scope of influencing factors compared to the more individualistic variables.

III. Data

The main point of my analysis is to look at the relationship between GDP per capita and income inequality, and in order to do this I will use the GINI coefficient as my dependent variable. The GINI coefficient is an index that measures income inequality within a nation and measures the inequality among values of a frequency distribution. The coefficients range from 0 to 1, with 0 representing a perfectly equal distribution, while 1 represents a perfectly inequal distribution. It is not possible for the value to be greater than 1 because that would suggest a negative income. The main reason that I am using the GINI index to measure income inequality is because it has the most widespread amount of data available for access online. It was also a common theme for economists to use this index within the literatures that I explored, so I will be following suite. The primary independent factor being analyzed is the log of GDP per capita. Although originally, I had used the regular GDP per capita, I then took the log of it instead which is common when looking at many different countries, and this led to the tests I conducted to become more substantial. In the hypothesis, it was stated that there would be a negative correlation between GDP per capita and income inequality, and the initial scatterplot shown in figure 1 suggests that this is true. Other variables which I will be accounting for in my regression models are unemployment, urban population, and savings. These variables and their descriptions can be found in Table 1.

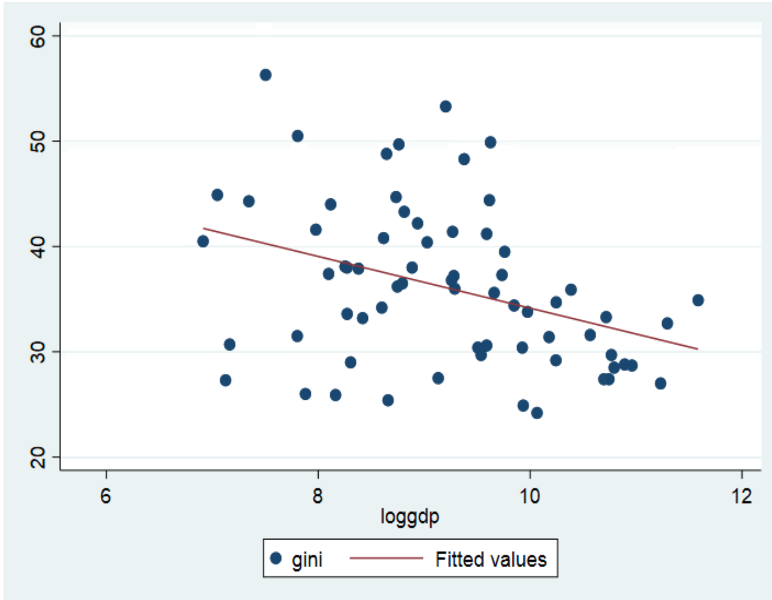

Figure 1

Table 1: Variables

Variable (unit)	**Description**	**Year**	**Source**
gini	Index measuring income inequality across population	2017	World Bank
loggdpcap (%)	Natural log of GDP per capita	2017	World Bank
unemp (%)	% of unemployed people in total labor force	2017	World Bank
urbpop (%)	% of urban population	2017	World Bank
gsav (%)	Gross savings, % of GDP	2017	World Bank

The variable table first lists the GINI coefficient, the dependent variable which was discussed earlier. Next is the main comparison that this work is trying to understand, which is the effect of GDP per capita on the income inequality. The hypothesis was that there is a negative correlation between the 2, meaning that if log GDP per capita increases, then the income inequality will decrease. There are also many other variables that are present which could affect a countries income inequality and GINI coefficient. This analysis will also be testing the unemployment rate as a % of the labor force. Previous research has shown that unemployment could have some influence over inequalities, so that is something that needs to be addressed. Another independent variable is the urban population as a percentage of population living in cities. When looking at the main topic of this paper, Kuznets curve will play a large role, and that curve emphasizes a shifting economy from agricultural to urban as a main influence of income inequality. Because of this, I want to analyze how much of the population is currently living in cities and where they are along the Kuznets Curve. The last independent variable I have selected is gross savings in US dollars. One of the main arguments for increased income inequality is that wealthy people are more likely to save and invest their money, therefore having a greater positive impact on the economy. This savings data will be used to see if this notion holds when compared to the GINI coefficient. Descriptive statistics for these variables can be found in table 2.

Table 2: Variable Statistics

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
gini	67	36.10	7.56	24.2	56.3
loggdpcap	251	8.81	1.41	5.68	12.06
unemp	233	7.27	5.01	0.14	27.04
urbpop	260	59.74	22.86	12.71	100
gsav	198	22.99	10.85	-48.78	68.39

Before going into the results, I want to look at and test the Gauss Markov assumptions.

1. **Linear in Parameters:**
   The model which was used from the book is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u,$ where y is the gini and $\beta_{0,} \ldots, \beta_k$ are the constants and u is the unobserved error. This equation does satisfy the first condition for being linear in parameters.

2. **Random Sampling:**

   In our models, we used data from every country in which the data was available on World Bank, including both high and low income countries. Therefore, our analysis was random, and the second condition is met.

3. **No Perfect Collinearity:**

   To test the collinearity of our models, STATA was used to test the correlations between the variables, and this can be found in appendix A. Although there was some correlation, the coefficients were not perfect (one to one), so the third assumption of no perfect collinearity is met.

4. **Zero Condition Mean:**

   The zero condition mean tests whether given the values of the independent variables, there is a 0 for the expected value of the error term u. Although this is hard to test for, looking at the residual plots in the simple regression, we can see that the expected error is not zero, and this forth condition is satisfied.

5. **Homoskedasticity:**

   This condition checks if the variance of error term u was kept constant throughout the regressions and independent variables. This can be seen in the figure 1 looking at the scatter plot and trend lines.

IV. Results

After testing all the Gauss Markov assumptions, we can now go over the results and interpretations of the models that were utilized to test the hypothesis. In particular, 3 models were utilized with regression analysis run on the countries that had variables, and the STATA outputs can be found in appendix B.

Model 1: Simple Regression:

The first model is a simple regression that looks at the relationship between income inequality (gini) and the GDP per capita. This can be written in the form of:

$$\text{gini} = \beta_0 + \beta_1(\text{loggdpcap}) + u$$

After running this regression on STATA, the output is:

$$\text{gini} = 58.685 - 2.454(\text{loggdpcap})$$

This first model has an R-squared of 0.137 which shows that there is a strong correlation between the gini and loggdpcap. Furthermore, there is a negative coefficient for lodgdpcap, supporting our hypothesis by showing that there is a negative relationship between the independent and dependent variable. This means that if the dependent variable, the gini, increases, then the loggdpcap will decrease by the amount of the coefficient.

Although the simple regression provides some useful information, there is only one variable being accounted for when there are many others which might have an impact on the gini, so it is not possible to conclude causality looking at the first model. This brings up omitted variable error, which will be dealt with in the next multiple regressions that are ran. These regressions which include the independent variables will give a better idea of the degree of the relationship between the two variables which are being researched.

Model 2:

The next regression will be a multilinear regression which includes all the independent variables which are being used to see the relation between the main two, gini and loggdpcap. In this model, the equation being utilized is:

$$\text{gini} = \beta_0 + \beta_1(\text{loggdpcap}) + \beta_2(\text{unemp}) + \beta_3(\text{urbpop}) + \beta_4(\text{gsav}) + u$$

In this model, the data set includes variables from every country in which the data is available, and adds the variables unemp, urbpop, and gsav to the simple regression which was run in model one. After the equation is put into STATA the output is show as:

$$\text{gini} = \beta_0 + -3.744(\text{loggdpcap}) - 0.042(\text{unemp}) + 0.186(\text{urbpop}) + -0.202(\text{gsav})$$

The R-squared increased from 0.137 in the first model to 0.246 in the second model after adding the other variables. This is commonly seen when adding more terms to a linear model and means that the relationship was strengthened further. Adding the new variables gives a better estimation of effects on income inequality and it can be seen that loggdpcap, unemployment, and government savings all had a negative coefficient, meaning they have an inverse relationship with the gini; as gini goes up, all of these other variables will decrease. On the other hand, urbpop has a positive coefficient, so when the percentage of people living in cities increases, the income inequality increases. This is surprising to see because intuitively, you would think that a country with more citizens living in cities would have less income inequality than those which are more agricultural. After running the second model, loggdpcap, the main explanatory variables coefficient increased from -2.454 to -3.744 meaning that there is an even bigger negative relationship between our main 2 variables. In terms of significance, loggdpcap was significant at the 1% level in both model 1 and 2. In the second model, urbpop was the only variable other than gdp that was significant at the 5% level, while unemp and gsav were both insignificant. Because of this, the third model will remove these variables and will only look at the loggdpcap and urban population.

Model 3:

For this model, unemp and gsav which both proved to be insignificant in model 2 were removed and this equation will be used:

$$\text{gini} = \beta_0 + \beta_1(\text{loggdpcap}) + \beta_2(\text{urbpop}) + u$$

Plugging this into STATA yields the output:

$$\text{gini} = \beta_0 - 4.568(\text{loggdpcap}) + 0.202(\text{urbpop})$$

This third model only contains 2 independent variables, but both the loggdpcap and urbpop's coefficients increased from the second model, as well as the R-squared which had a slight increase to 0.267. With each model, the R-squared increased to some degree which shows that the data fits the regression model with good fit. The coefficient of loggdpcap once again increased to -4.568, showing a very strong negative correlation to the gini index. There was also an increase in gsav from 0.168 to 0.202, showing a relatively strong positive relation. This term also stayed significant at the 5% value which is a good sign. The positive value of gsav's coefficient is still confusing looking at the Kuznets Curve which states that a country that transitions from an agricultural to an urban population sees a decrease in income inequality. However, some research such as the one I included in the literature review shows that it is possible for some countries to experience more inequality as the transition is made. Overall, this model shows how big of a negative impact GDP per capita has on income inequality, supporting our initial hypothesis strongly.

Table 3 below shows the summary of all the regression models which have been explained above.

Table 3: Estimation Results

Independent Variables	**Model 1**	**Model 2**	**Model 3**
loggdpcap	-2.454***	-3.744***	-4.568***

	(0.765)	(1.052)	(0.954)
unemp		-0.042 (0.178)	
urbpop		0.168** (0.068)	0.202** (0.063)
gsav		-0.202 (0.134)	
Intercept	58.68*** (7.094)	63.92*** (7.561)	64.64*** (6.812)
Number of observations	67	63	66
R - Squared	0.137	0.246	0.267

V. Extensions

Now that the regressions and models have been made, we can look at significance using the F-test. We saw in model 2 that both unemp and gsav were individually insignificant, but we can now use the F-test instead of the T-test to see the joint significance of these variables. The unrestricted model will be 2 because it included all the variables which were being looked at, and model 3 will be utilized as the restricted. Looking at model 2 and the variables unemp and gsav, we can hypothesize that:

$$H_0: \beta_2 = 0, \beta_4 = 0$$

$$H_1 : H_0 \text{ is false}$$

To calculate the F-statistic, we first divide the difference of residuals between the restricted and unrestricted models and divide it by 2. Next, the residual of the unrestricted model is divided by the degrees of freedom. After dividing these steps, it is concluded that the F-stat for unemp and gsav is

2.98. The critical value at 10% is 2.40, so the null hypothesis is rejected, meaning unemp and gsav are jointly insignificant.

Model 4:

In the last model, the poverty headcount of countries was included, with the data showing what percentage of the population lives under a certain monetary value. This was plugged into the third model. After doing so, the output STATA gave was:

$$\text{gini} = \beta_0 - 3.333(\text{loggdpcap}) + 0.210(\text{urbpop}) + 0.299(\text{povhead})$$

It was surprising to see that after adding this variable, loggdpcap stayed significant at the 1% level, and the urbpop became significant at 1%. The poverty headcount showed a relatively strong correlation with the gini while being significant at the 5% level.

VI. Conclusion

Looking at the 4 different models that were utilized, all of them supported the initial hypothesis that gini and GDP per capita have a negative relation with one another. In the first simple regression model, loggdpcap had a coefficient of -2.454 showing a very strong negative correlation, while having an R-squared of 0.137. This could also be seen in the scatter plot using the line of best fit. This is not really surprising looking from an intuitive perspective because you would expect that if GDP for areas are high, then there would be low levels of income inequality. For each model that was ran, it was a good sign to see that the R-squared values increased, showing stronger levels of relations as variables were added and taken out accordingly.

For the second model that was ran, it was surprising to see that both unemployment and gross savings were insignificant. Unemployment is complex in many countries, but it is normally a trend that countries with high unemployment will face higher levels of poverty and income inequality. When looking at gross savings, those with higher levels of income are more likely to partake in saving because they have disposable money. Those who are living in poverty are less likely to invest or save because they live day to day and can not afford to do so. Further studies could be conducted to see why these are insignificant. An interesting observation made in the analysis is that urbpop had a positive correlation with the gini index. Most would expect that a country that has more of its citizens living in the cities would experience less income inequality, but the positive coefficient suggests the opposite. Many literatures have been written on the Kuznets Curve, which states that a countries income inequality will increase as a country is transiting to becoming more urbanized, but will then decrease as the transformation is complete, leading to a U-shaped curve. This data could indicate that there are currently many countries making the transition, causing the percentage of people living in urbans population to have a positive relation to the gini. Overall, it can be concluded with confidence that the hypothesis was supported, and that GDP per capita and income inequality have a negative relation to one another. To further expand on this research, others could include more variables, while also analyzing their effects.

Appendix A: Correlation outputs

```
. pwcorr gini loggdpcap Unemployment Urbanpopulation GrossSavings
```

	gini	loggdp~p	Unempl~t	Urbanp~n	GrossS~s
gini	1.0000				
loggdpcap	-0.3696	1.0000			
Unemployment	0.1269	0.0391	1.0000		
Urbanpopul~n	0.0022	0.7105	0.0975	1.0000	
GrossSavings	-0.2845	0.2657	-0.1490	0.1442	1.0000

Appendix B: Stata Outputs

Model 1:

```
. regress gini loggdpcap
```

Source	SS	df	MS
Model	514.662719	1	514.662719
Residual	3252.77668	65	50.0427182
Total	3767.4394	66	57.0824152

Number of obs	=	67
F(1, 65)	=	10.28
Prob > F	=	0.0021
R-squared	=	0.1366
Adj R-squared	=	0.1233
Root MSE	=	7.0741

gini	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
loggdpcap	-2.453983	.7652101	-3.21	0.002	-3.982213	-.9257532
_cons	58.68464	7.094332	8.27	0.000	44.51628	72.853

Model 2:

```
. regress gini loggdpcap Unemployment Urbanpopulation GrossSavings

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(4, 58)        =      4.73
       Model |  806.570947         4  201.642737   Prob > F        =    0.0023
    Residual |  2470.68334        58  42.5979886   R-squared       =    0.2461
-------------+----------------------------------   Adj R-squared   =    0.1941
       Total |  3277.25429        62  52.8589401   Root MSE        =    6.5267

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   loggdpcap |  -3.743802   1.052484    -3.56   0.001    -5.850577   -1.637027
Unemployment |  -.0417712   .1776414    -0.24   0.815    -.3973591    .3138166
Urbanpopula~n |   .1683885   .0677125     2.49   0.016     .0328472    .3039297
GrossSavings |  -.2022611   .1341086    -1.51   0.137    -.4707086    .0661863
       _cons |   63.92156   7.560729     8.45   0.000     48.78712      79.056
------------------------------------------------------------------------------
```

Model 3:

```
. regress gini loggdpcap Urbanpopulation

      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(2, 63)        =     11.46
       Model |  991.234147         2  495.617073   Prob > F        =    0.0001
    Residual |  2724.98843        63  43.2537846   R-squared       =    0.2667
-------------+----------------------------------   Adj R-squared   =    0.2435
       Total |  3716.22258        65   57.172655   Root MSE        =    6.5768

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   loggdpcap |  -4.568034   .9542387    -4.79   0.000    -6.474928    -2.66114
Urbanpopula~n |   .2015339   .0632542     3.19   0.002     .0751304    .3279374
       _cons |   64.63866   6.812432     9.49   0.000      51.0251    78.25222
------------------------------------------------------------------------------
```

Model 4:

```
. regress gini loggdpcap Urbanpopulation Povertyheadcount

      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(3, 62)        =     11.44
       Model |  1324.34498         3  441.448327   Prob > F        =    0.0000
    Residual |  2391.87759        62  38.5786709   R-squared       =    0.3564
-------------+----------------------------------   Adj R-squared   =    0.3252
       Total |  3716.22258        65   57.172655   Root MSE        =    6.2112

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   loggdpcap |  -3.333262    .994348    -3.35   0.001    -5.320936   -1.345589
Urbanpopula~n |   .2099894   .0598073     3.51   0.001     .0904363    .3295426
Povertyhead~t |   .2986808   .1016451     2.94   0.005     .0954951    .5018665
       _cons |   51.52971    7.82911     6.58   0.000     35.87954    67.17987
------------------------------------------------------------------------------
```

References

Abhijit V. Banerjee, and Esther Duflo. "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth (Boston, Mass.)* 8.3 (2003): 267-99.

Barro, Robert J. (2000). Inequality and Growth in a Panel of Countries. Journal of Economic Growth, 5, 5-32.

Ikemoto, Yukio, and Uehara, Mine. "Income Inequality and Kuznets Hypothesis in Thailand." *Asian Economic Journal* 14.4 (2000): 421-43. Web.

Kuznets, Simon. “Economic Growth and Income Inequality.” The American Economic Review, vol.45, no. 1, 1955, pp. 1–28., www.jstor.org/stable/1811581.